DOCUMENT RESUME

ED 078 005 TM 002 817

AUTHOR Bruno, Nancy L.
TITLE Implementation Procedures for Statewide Assessment.
PUB DATE Feb 73
NOTE 15p.; Paper presented at National Council for Measurement in Education Meeting (New Orleans, Louisiana, February 26-28, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Analysis; Educational Objectives; *Evaluation Methods; Evaluation Techniques; *Measurement Techniques; *Program Evaluation; Speeches; *State Programs; Student Evaluation

ABSTRACT

Six principles for statewide assessment are discussed: (1) involve the community; (2) specify and define goals; (3) use measuring devices with face and content validity; (4) take noncognitive effects of school into account; (5) design data presentation for lay understanding; and (6) do not let assessment be an end in itself. Decisions to be made in planning and conducting a statewide assessment program are also discussed. They involve: goal setting, establishing priorities, the number of goals to be assessed, the target population, sampling procedures, instrumentation, correlates of achievement, data analysis, reporting of results, and when to conduct the assessment. (KM)

ED 078005

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 002 817

IMPLEMENTATION PROCEDURES FOR STATEWIDE ASSESSMENT

Nancy L. Bruno

Center for Statewide Educational Assessment

Educational Testing Service

Paper Presented at
NCME
February, 1973

FILMED FROM BEST AVAILABLE COPY

PRINCIPLES OF ASSESSMENT

Although the title of this paper is listed in the program as "Implementation Procedures for Statewide Assessment" a more appropriate title probably is "Principles, Pitfalls and Strategies in Statewide Assessment." The personal experiences of the staff at the Center for Statewide Educational Assessment and our knowledge of various state programs led to the formulation of six assessment principles. These were designed as a guide for state department personnel and others to assist them in optimizing their chances for a successful assessment program. Although they have been published elsewhere, I feel so strongly about their underlying importance in any assessment program, I am going to repeat them here.

Involve the community. Effective educational assessment demands the recognition and involvement of the entire community, i.e., legislators, educators, parents, students, business managers, labor leaders and other concerned groups.

Although this can be done at several points in the program, ideally the community should be involved to some degree at the very beginning.

One method of involving them is to have representatives from each group assist in determining what the goals for education ought to be. Since each group may have different priorities this could be a time consuming activity. The time will be well spent, however, since in addition to determining the goals, the participants should also become aware of each other's needs and constraints.

For example, the legislator primarily wants to know how much pupil learning and development the money he appropriates for education is buying.

He must also answer to his constituents who may not reelect him if they feel he is not concerned about the quality of their children's education.

Teachers have a great interest in assessment since they are directly concerned with education. Some may have negative attitudes toward it if they feel they personally will be evaluated solely on the basis of assessment results for their students. In addition to the valuable contribution they can make, they will be less apt to feel threatened because they have been given the opportunity to participate in the developmental phases of the program.

Students should certainly have representation in deciding what the goals for education ought to be, since those goals most directly concern them and their future.

Parents want assurance that their children are receiving the kind of education that will enable them to cope with the ever increasing complexity of the world in which they live.

The amount and type of community involvement in the assessment program will depend, to a great extent, on the time constraints within which the program must operate. If there is sufficient time, it would be advisable to have a series of regional meetings with representative groups. If pressed for time, the most efficient way to involve the community is to select an advisory committee. The members of the committee should be selected so that all the concerned groups are represented and allowed to contribute. If they are used merely as a rubber stamp for a fait accompli, the purposes of community involvement are defeated.

Meaningful early involvement of the various interest groups should facilitate understanding and cooperation when the assessment is conducted.

Specify and define goals. After the broad goals have been identified and accepted, they must be defined operationally and behaviorally so they can be measured. The community should continue to be consulted during this phase; especially the educators.

An example of the need for this type of definition is the goal "To appreciate human endeavor in the arts." As stated, it is too broad to measure. To illustrate, one facet of this goal might be to demonstrate an appreciation of music. An appreciation of music could be defined behaviorally as the number of times tapes, records and music books are used. This definition corresponds to the receiving and responding levels in the taxonomy of the affective domain (Krathwohl, et al, 1964). The behavioral objective could then be measured by a frequency count of the tapes, records and music books used in the library and those taken off campus for listening and reading. The number of usages and the proportion of students involved would be an indicator of the student body's appreciation of music.

Measuring devices must have face and content validity. The instruments used in the assessment program, whether selected from existing tests or constructed specifically for the program, should contain an adequate sampling of the specified universe of content. In addition, they should be face valid, i.e., the layman must be able to look at the instruments and see the relationships between them and the goals being measured. For example, if the objective is to measure understanding and the instrument contains items that are purely factual in content, the instrument would not have content validity although it might appear to be face valid. Adequate assessment devices must present both.

Take noncognitive effects of school into account. Society, for many reasons, is delegating more and more responsibility to the schools for

developing learning outcomes which are not skills centered. The appreciation of human endeavor in the arts mentioned earlier is one example. Another is the development of a positive self concept. Although these noncognitive areas are admittedly more difficult to measure and interpret, they must not be ignored in the early phase of an assessment program or they most likely will continue to be neglected as the program is enlarged. A second reason for including noncognitive measures early is that people tend to concentrate efforts on the areas being evaluated. Therefore, failure to evaluate noncognitive areas has the effect of focusing the educational process on the skill development segment of education to the neglect of the equally important noncognitive areas.

Data presentation should be designed for lay understanding. Possibly the most crucial aspect in determining the success or failure of an assessment program is the reporting of results. The reports should be in terms that are comprehensible to the layman. Interpretation of statistical data, particularly that which requires qualification, such as test scores, is most effective when interaction between the receiver and presenter is possible. However, there is likely to be little interaction if the results are reported in sophisticated technical terms. Possible alternatives for use in the presentation of data are: expectancy tables based on previous year's performance; comparison with state norms; percentage of response to each option of the key items; description of the distribution of student scores in terms of the kinds of items which they can handle successfully and those which present difficulty; and in relation to attainment of the goals. The method selected for reporting results will depend on several factors. Among these are the uses which will be made of the assessment data, who will use the results and the type of instruments used.

Assessment must not be an end in itself. The last principle, which perhaps should have been first, is that assessment must be clearly identified as one component of the total educational process. Evaluative data are collected to meet specific needs and if assessment is not related to these purposes it is useless. Assessment must provide decision makers at various levels with information that will enable them to make program modifications necessary for educational improvement. For example, high and low scoring schools should be observed to determine the activities, materials, etc. that seem to be making a difference in the student output so that these may be tried out in other schools.

PITFALLS AND STRATEGIES

Assessment, like education itself, is a dynamic process and may be initiated from different points depending on the constraints under which assessment is conducted. If a state has a legislative mandate to assess specified areas for instance, it would enter the system at a different point than one that had no constraints. A state with formalized educational goals would begin at a different point than one with only implicit goals.

Before implementing an assessment program there are many questions that must be answered. In this paper assessment decisions are treated in a linear fashion although in reality they are nonlinear and interdependent.

The "umbrella" question for the entire assessment program is "Why assess?" The most defensible reason for conducting an assessment is to determine the status of education thereby providing decision makers with information that will allow them to examine programs that are succeeding to determine their generalizability and applicability to other locations and student populations.

Although there are educators and others who say that it is not necessary to determine explicitly stated goals before assessment, in several states where assessments have been conducted without them, there has been a great deal of resistance to the program. On the other hand, programs in some states that have devoted a great deal of time to determining the goals and then assessed only the basic skills, have also met resistance and even hostility. It may be that having or not having goal statements had nothing to do with the success or failure of these programs, but were instead

indicative of poor communication among the various groups concerned with assessment. In any event, it would seem that the process of community involvement in deriving goal statements would enhance a program's chances for success by eliminating unknown factors.

Once there is a set of formal goals, the next step is to determine their priorities. If the community was not involved formally in stating the goals, it should certainly be involved at this stage--especially the educators. This can be accomplished in several ways. Provided there is sufficient time, it is best to conduct a series of regional meetings with representatives of the various interest groups. A second alternative is to select a cross-sectional advisory committee whose members represent the different publics concerned with and affected by assessment. The Delphi technique, originally used in military forecasting, is a means of involving the community without the personal contacts that occur in regional meetings. Georgia is one state that has used this technique with apparent success. Each participant checkrates lists of statements and writes his comments. Successive mailings with the results of the previous evaluation are used to reach consensus.

After priorities have been established, decisions must be made with regard to how many of the goals will be assessed. Financial constraints may limit the number, but it is recommended that at least one noncognitive goal be included in the first assessment. Attitude toward learning and attitude toward self would be likely candidates for inclusion since they are related to the cognitive areas and there are measuring devices available.

The question of who will be assessed has two facets; the grades or ages of those to be tested and whether there will be universal testing or sampling.

Once again these decisions depend on others. For example, if National Assessment exercises are chosen, they are only appropriate for certain ages or grades. In general, states with operational assessment programs are assessing at least one elementary and one secondary grade. This decision is usually arbitrary or expedient. The elementary grades 3, 4 and 5 are frequently selected. Some of the reasons for choosing those grades are: it is the year no other tests are given; it is the year a test that is part of the assessment package is given; tests are available for that grade; or there is time to make program changes that will benefit these children. Testing at the secondary level ranges from grade 7 to grade 12. At this level the problem of dropout and college admissions testing may influence the grade selection. The selection is also influenced by whether the data will be used to improve programs before the students leave school or to assess the overall performance of the schools.

Following the selection of the population, the decision must be made whether to test all students or to sample. If the results are to be used for diagnostic purposes, all students should be tested and instruments must be long enough to provide reliable results within the diagnostic categories. To provide a statewide picture of education a sample is sufficient. Because of the complexity of sampling procedures, it is recommended that a sampling expert be consulted before making the final decision. Once the lower unit price for the increased volume in testing all students and the cost of drawing a sample are computed, it may be financially as economical to test everyone in a given grade as to sample.

The political climate may be such that the program would suffer a loss in credibility if sampling were used. For example, even though politicians rely on sample based polls to predict their elections, they never trust the

"other guy's" poll. Other problems with sampling are the increased difficulty in interpreting results and the difficulty in explaining to parents how the results are representative of their school district or building when their child was not included in the sample.

Even when sampling is used, large amounts of student time are still required for the assessment. One way to decrease the amount of student time is to use matrix sampling in which both students and items are sampled. In considering this technique however, the increased costs for printing, administration and analysis must be weighed against the time saved to determine its feasibility. It should be reiterated that matrix sampling may allow district or building descriptions, but for any unit as small as a classroom, sampling will probably not produce sufficiently reliable results.

The next problem is the selection of measuring devices. The alternatives here are many and the choice depends on the reasons for the assessment and the uses to be made of the results.

When one of the purposes is to compare the state to a regional or national group, some type of test for which there is normative data must be used. Both standardized achievement tests and the National Assessment exercises permit these kinds of comparisons.

If there is no requirement to make comparisons additional possibilities are: (1) construct new instruments locally; (2) contract with a test publisher to construct instruments to specifications; (3) aggregate the data from tests routinely administered by school districts. This last alternative will become more feasible when the Office of Education releases the data from the National Test-Equating Study in Reading. This study equated scores for the seven reading comprehension and vocabulary tests most widely used for children in grades 4, 5 and 6.

When intrastate comparisons are to be made from assessment results it is essential to collect information concerning the conditions of learning that may be related to achievement. It is uninformative and misleading to compare institutions from a wealthy surburban district with those from a poor rural or inner-city district without considering the variables that have been found to be associated with achievement. If compared solely on the basis of test results, the poorer districts would appear to be doing less for their students. However, when student background characteristics, teacher characteristics and financial resources available are also considered, the poorer districts could be making a more efficient use of their resources to improve student performance than the wealthier districts.

Probably the most important reason for examining condition variables is to provide areas for hypothesis generation about the causes of learning success or failure. Socioeconomic status variables have consistently been found to be related to school achievement. This does not mean that being poor or wealthy determines a student's achievement, but it does indicate that there may be experiences available in communities of differing SES strata that account for the differing achievement results. Preliminary reports from a state assessment follow up study of high and low scoring schools indicate that teachers tend to interact the same with all groups when it is doubtful that the same approach is effective for all groups. In addition, unless the achievement related variables are examined, there is no way of knowing whether previously found relationships are true for a particular state or district in that state.

A paper by Campbell (in press) reviews the uses of correlates of achievement and presents some procedures for examining what actually occurs in classrooms that might account for differences on achievement measures.

Although these procedures could probably not be used on a statewide basis because of the cost and time, the state could provide financial and technical assistance to local districts to carry out intensive local assessments. The state could also serve as a clearinghouse to distribute the findings from the local studies.

Once all of the data are collected, they must be analyzed. The kinds of analyses will be determined primarily by the types of data collected and the purposes of the assessment. Frequency distributions, a measure of central tendency and a measure of variability will provide a description of the position of the schools in the state and permit comparison with a norm group.

When data on condition variables have been collected, but not quantified, meaningful analyses are the two-way factorial analysis of variance and the Friedman two-way analysis for ranked data. These analyses reveal interrelationships which should be examined further.

If the data collected on condition variables are quantifiable, a factor analysis will determine clusters that may yield meaningful interpretations in terms of educational implications.

Multiple correlation procedures allow more complex relationships to be considered. They provide a method for examining the unique contribution of many variables in a systematic way. Again, it must be remembered that results obtained from these analyses do not indicate any causation. From these results, variables can be identified which should be studied further to determine possible influences on students' learning experiences. The next step is to conduct intensive examination of specific learning environments and programs to determine variables that may be making a difference. The intensive studies of this type should provide information which will enable changes to be made.

The results of these changes in the program or environment can then be evaluated.

Perhaps the most critical phase of an assessment program is the reporting of results. Interpretation of statistical data is probably the most difficult part of an assessment program for many educators and for most of the community at large.

The ideal method of reporting results is where interaction between the receiver and presenter is possible. This personal contact is especially important when assessment results are reported to legislators, state board of education, the governor's office and other decision makers. If these people are only presented with a written report containing masses of data, chances are it will never be used (or it will be misused) in making educational decisions. This is not to say that there should not be a written report. However, the written report should be used as a reference after there has been an opportunity for discussion of the results.

The written report might contain expectancy tables; comparison with national, state and regional norms; percentage of response to each option of key items that reflect the concept being measured and a description of the distribution of children in terms of the kinds of problems with which they can deal successfully and those with which they cannot.

Ideally the reports would be interpreted for each school district at a local meeting with administrators, teachers and perhaps some parents. If this is not possible because of financial or time constraints, the next best alternative is to conduct regional meetings with two or three representatives from each district. The district representatives would then report to their districts.

Whatever method is used to report results, part of the report should be devoted to a discussion of the implications and uses of the data. To tell districts that their students are performing poorly without having staff available to respond to requests for assistance serves no educationally useful purpose. Hopefully, the assessment program has not been conducted in a vacuum and at this point guidance personnel, curriculum specialists and researchers would be prepared to assist those local districts requesting help.

The operational details of the assessment need to be carefully outlined with a clear delineation of responsibilities and critical dates. PERT diagrams are one way of keeping track of the "nitty gritties" of the program. There are also computer programs which will provide such things as critical dates and manpower needed to keep the assessment activities on schedule.

The last decision to be considered in this paper is when to conduct the assessment. The uses of the data and the amount of time needed to report the results back to the local school districts are two of the parameters that must be considered before determining when the assessment is conducted.

Generally, early fall testing is recommended so that results can be reported and studied in time for decision makers, at both the state and local levels, to make use of them in budgetary planning. This recommendation should not be misinterpreted as implying that assessment results should be used as a basis for rewarding or punishing local districts. Rather they should be used to help determine possible areas for change that may or may not require additional funds. For example, if a superintendent discovered from the assessment results that two schools in his district were doing

very well in reading and a third was doing very poorly, he might decide to have his reading specialist spend more time at the school in which students performed poorly. If he had no reading specialist, he might decide to reallocate some of his funds to hire one. At the state level, additional funds might be specifically allocated for hiring a reading specialist.

In summary, after a discussion of six principles of assessment, I indicated the decisions and some alternative strategies for planning and conducting a statewide assessment program. These decisions include goal setting, establishing priorities, the number of goals to be assessed, the target population, sampling procedures, instrumentation, correlates of achievement, data analysis, reporting of results and when to conduct the assessment.

The final decisions in assessment must be state specific since it is highly unlikely that any two states operate under the same constraints with the same parameters.

REFERENCES

Bruno, N. L., Campbell, P. B., and Schabacker, W. H. Statewide Assessment: Methods and Concerns. Princeton, N. J.: Center for Statewide Educational Assessment, Educational Testing Service, 1972.

Campbell, P. B. The Use of Correlates of Achievement in Statewide Assessment. Princeton, N. J.: Center for Statewide Educational Assessment, Educational Testing Service, (in press).

Krathwohl, D. R., Bloom, B. S., and Masia, B. B. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, 1956.